

# Zadanie rekrutacyjne, DSC PJATK 2025

Witamy w zimowej edycji konkursu rekrutacyjnego organizowanego przez Data Science Club (DSC) przy PJATK w Warszawie. Ta edycja opiera się na syntetycznym zbiorze danych o konsumpcji treści audio i pozwala sprawdzić umiejętności modelowania, analizy danych oraz pracy zespołowej.

---

## Harmonogram

- Start konkursu: 14.10.2025
  - Deadline zgłoszeń dziennych: 23:50, 28.10.2025
  - Ogłoszenie wyników: 23:55, 28.10.2025
  - Ocena zgłoszeń: 01.11.2025
- 

## Cel Konkursu

Sprawdzamy, jak uczestnicy radzą sobie z analizą danych, budową modeli prognostycznych oraz pracą zespołową w realistycznym scenariuszu. Konkurs dzieli się na dwie główne części:

1. Zbudowanie modelu przewidującego, jak długo użytkownik będzie słuchał odcinka audio.
  2. Przygotowanie prezentacji (bądź notebooka) podsumowującej proces analizy danych, modelowania oraz wyciągniętych wniosków.
- 

## Wymagania

1. Rozwiązanie Kaggle Competition
  - Każdy zespół ma za zadanie opracować model przewidujący, jak długo użytkownik będzie słuchał odcinka audio.
  - Szczegóły zasady oceniania oraz informacje o zbiorze danych, znajdują się na stronie konkursu: [Kaggle Competition](#).
2. Repozytorium GitHub
  - Należy utworzyć repozytorium na GitHub zawierające cały kod, opis rozwiązania i prezentację.
  - Kod powinien być czytelny, dobrze udokumentowany i zgodny z dobrymi praktykami.
  - Mile widziane są notatniki Jupyter z analizą oraz komentarzami.

### 3. Prezentacja wyników

- Na koniec zadania każda drużyna musi przygotować prezentację zawierającą:
  1. Minimum 2 wizualizacje z danych, które opowiadają jakąś historię,
  2. Opis procesu analizy i modelowania,
  3. Największe wyzwania i jak sobie z nimi poradziście.
- Prezentacja może zostać przygotowana w formie klasycznej (np. w PowerPoint) lub jako notatnik Jupyter Notebook.
- Jeśli prezentacja jest w formacie Jupyter Notebook, powinna być podzielona na dwa osobne pliki:
  1. EDA + wnioski (Eksploracyjna analiza danych i kluczowe obserwacje)
  2. Modelowanie + wnioski (Opis użytych metod modelowania i ich efektywność)

### 4. Zgłoszenie mailowe

- Każdy zespół musi przesłać do 26 października 2025, najpóźniej do godziny 23:59 maila na [dsc@pjwstk.edu.pl](mailto:dsc@pjwstk.edu.pl) z następującymi informacjami:
  1. Link do repozytorium GitHub
  2. Numer grupy i skład zespołu
  3. Załączoną prezentację (w przypadku prezentacji w formacie PDF lub PPTX)

## Ocena zespołów

- 3 najlepsze zespoły (zgodnie z wynikami na Kaggle) dostają się do DSC PJATK automatycznie.
- Zespoły poniżej 3. miejsca będą oceniane na podstawie:
  - Wyników modelu
  - Jakości kodu i organizacji repozytorium
  - Kreatywności w podejściu do problemu
  - Użyteczności i jakości przetwarzania danych

## Zasady fair play

1. Możecie korzystać z dowolnych źródeł i narzędzi do pracy z danymi i modelem.
2. Nie wolno kopiować kodu od innych zespołów – w przypadku uzyskania dostępu do ich kodu, nie możecie się nim “inspirować” 1 do 1.
3. Wszelkie podejrzenia o plagiat będą skutkowały dyskwalifikacją zespołu.
4. Każda drużyna ma ustalony limit pięciu dziennych zgłoszeń, każda drużyna musi się do niego stosować. Wszelkie próby obejścia limitu (np. tworzenie wielu kont) będą skutkować dyskwalifikacją.
5. Wszelkie próby sztucznego podbijania oceny, takie jak używanie bezpośrednio ukrytych wartości testowych, mogą skutkować obniżeniem oceny lub całkowitym pominięciem wyników zespołu.

## Motywacja dla uczestników

To zadanie to szansa na stworzenie wartościowego case study do portfolio. Nawet mniej skuteczna próba przyniesie nowe wnioski i doświadczenie. Podchodźcie kreatywnie, testujcie hipotezy i bawcie się danymi. Powodzenia i do zobaczenia na leaderboardze!